

Chromatin analysis in human early development reveals epigenetic transition during ZGA

Jingyi Wu^{1,7,8,9}, Jiawei Xu^{2,9}, Bofeng Liu^{1,9}, Guidong Yao^{2,9}, Peizhe Wang^{3,9}, Zili Lin^{1,9}, Bo Huang⁴, Xuepeng Wang^{5,6}, Tong Li², Senlin Shi², Nan Zhang², Fuyu Duan³, Jia Ming³, Xiangyang Zhang², Wenbin Niu², Wenyan Song², Haixia Jin², Yihong Guo², Shanjun Dai², Linli Hu², Lanlan Fang², Qiuqun Wang¹, Yuanyuan Li¹, Wei Li^{5,6}, Jie Na^{3*}, Wei Xie^{1*} & Yingpu Sun^{2*}

Upon fertilization, drastic chromatin reorganization occurs during preimplantation development¹. However, the global chromatin landscape and its molecular dynamics in this period remain largely unexplored in humans. Here we investigate chromatin states in human preimplantation development using an improved assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)². We find widespread accessible chromatin regions in early human embryos that overlap extensively with putative *cis*-regulatory sequences and transposable elements. Integrative analyses show both conservation and divergence in regulatory circuitry between human and mouse early development, and between human pluripotency *in vivo* and human embryonic stem cells. In addition, we find widespread open chromatin regions before zygotic genome activation (ZGA). The accessible chromatin loci are readily found at CpG-rich promoters. Unexpectedly, many others reside in distal regions that overlap with DNA hypomethylated domains in human oocytes and are enriched for transcription factor-binding sites. A large portion of these regions then become inaccessible after ZGA in a transcription-dependent manner. Notably, such extensive chromatin reorganization during ZGA is conserved in mice and correlates with the reprogramming of the non-canonical histone mark H3K4me3, which is uniquely linked to genome silencing^{3–5}. Taken together, these data not only reveal a conserved principle that underlies the chromatin transition during mammalian ZGA, but also help to advance our understanding of epigenetic reprogramming during human early development and *in vitro* fertilization.

We sought to investigate accessible chromatin in human preimplantation embryos using ATAC-seq². Our optimized ATAC-seq protocol (miniATAC-seq; Methods) can acquire high-quality data using as few as 20 cells (Extended Data Fig. 1a–c). We collected two-cell and eight-cell human embryos, inner cell masses (ICMs) from embryonic day 5 (E5) human blastocysts, and human embryonic stem (ES) cells (Extended Data Fig. 2a). The isolated ICMs are post lineage segregation and probably include epiblast and primitive endoderm cells (Extended Data Fig. 2b). We also performed RNA sequencing (RNA-seq) analysis for each stage and validated these data (Extended Data Fig. 2c, d). We then conducted two replicates of ATAC-seq for each stage, and an additional two replicates using triploid embryos (derived from zygotes with three pronuclei, or 3PN), which are more available during *in vitro* fertilization (IVF). All replicates showed highly consistent results (Fig. 1a, Extended Data Fig. 3a, b and Supplementary Table 1). We identified 22,977, 40,426, 44,907 and 46,172 ATAC-seq peaks for the two-cell embryos, eight-cell embryos, ICMs and ES cells, respectively, with comparable genome coverages to that for ES cell DNase sequencing (DNase-seq; Extended Data Fig. 3c). ICM ATAC-seq peaks

have the highest overlap with ES cell DNase I hypersensitive sites, and the overlap decreased in early-stage embryos (Extended Data Fig. 3d). *NANOG* is induced at the eight-cell stage, which coincides with ATAC-seq enrichment at promoters and nearby putative enhancers (Fig. 1b). *POU5F1* is slightly upregulated at the eight-cell stage and is highly activated in the ICMs and ES cells (Fig. 1b). Notably, ATAC-seq peaks near *POU5F1* transit from intronic regions to promoter upstream loci from eight-cell human embryos to ES cells, indicating a likely switch of regulatory elements between naive and primed pluripotency, as in the mouse⁶. Concordant gene expression and ATAC-seq was also found for *ZSCAN5B* (Extended Data Fig. 3e). Genome-wide, as human major ZGA starts between the four- and eight-cell stages⁷, we identified stage-specifically expressed genes and examined their promoter accessibility (Extended Data Fig. 4a). We found promoters with constantly high accessibility that are preferentially CpG-rich as reported previously⁸, whereas promoters with constantly low accessibility are generally CpG-poor. Interestingly, a group of genes showed dynamic promoter accessibility that correlates with gene expression. These genes preferentially function in development, differentiation and morphogenesis, as also evidenced by their histone 3 Lys27 trimethylation (H3K27me3) enrichment in ES cells and fibroblasts (Extended Data Fig. 4b, c). Therefore, promoter accessibility in early human development correlates with both gene activities and CpG densities (Extended Data Fig. 4d).

Next, we sought to investigate distal open chromatin loci. Certain repeats are highly active in mammalian preimplantation embryos⁹. In contrast to ES cells and somatic cells, distal ATAC-seq peaks preferentially enrich for repeats at the eight-cell stage, especially for Alu, SINE and long-terminal repeat (LTR) retrotransposons (Extended Data Fig. 5a). ERVK, SVA and ERV1 are enriched at both the eight-cell and ICM stages. Both ERVK and SVA are expressed in human preimplantation embryos⁹. Thus, accessible chromatin is extensively shaped by transposable elements upon human ZGA. As distal accessible chromatin frequently harbours transcription factor-binding sites, we attempted to infer key factors that may regulate transcription circuitry in early development. We first identified stage-specific distal ATAC-seq peaks (Extended Data Fig. 5b). Eight-cell specific peaks are generally present near genes involved in RNA localization and processing. By contrast, distal peaks in ES cells, but not those in the ICM, are preferentially associated with genes involved in SMAD/activin signalling and stem-cell regulation. Notably, activin A is a core signalling module that maintains primed ES cells but destabilizes naive pluripotency¹⁰. Interestingly, we also observed widespread open chromatin in the pre-ZGA two-cell embryos (Extended Data Fig. 5b). However, unlike in other stages, their predicted targets (Methods) generally lack transcripts at the two-cell stage. It is possible that pre-ZGA accessible elements bind transcription

¹Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, THU-PKU Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China.

²Center for Reproductive Medicine, Henan Key Laboratory of Reproduction and Genetics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China. ³Center for Stem Cell Biology and Regenerative Medicine, School of Medicine, Tsinghua University, Beijing, China. ⁴PKU-THU Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ⁵State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. ⁶University of Chinese Academy of Sciences, Beijing, China.

⁷Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ⁸Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁹These authors contributed equally: Jingyi Wu, Jiawei Xu, Bofeng Liu, Guidong Yao, Peizhe Wang, Zili Lin. *e-mail: jie.na@tsinghua.edu.cn; xiewei121@tsinghua.edu.cn; syp2008@vip.sina.com

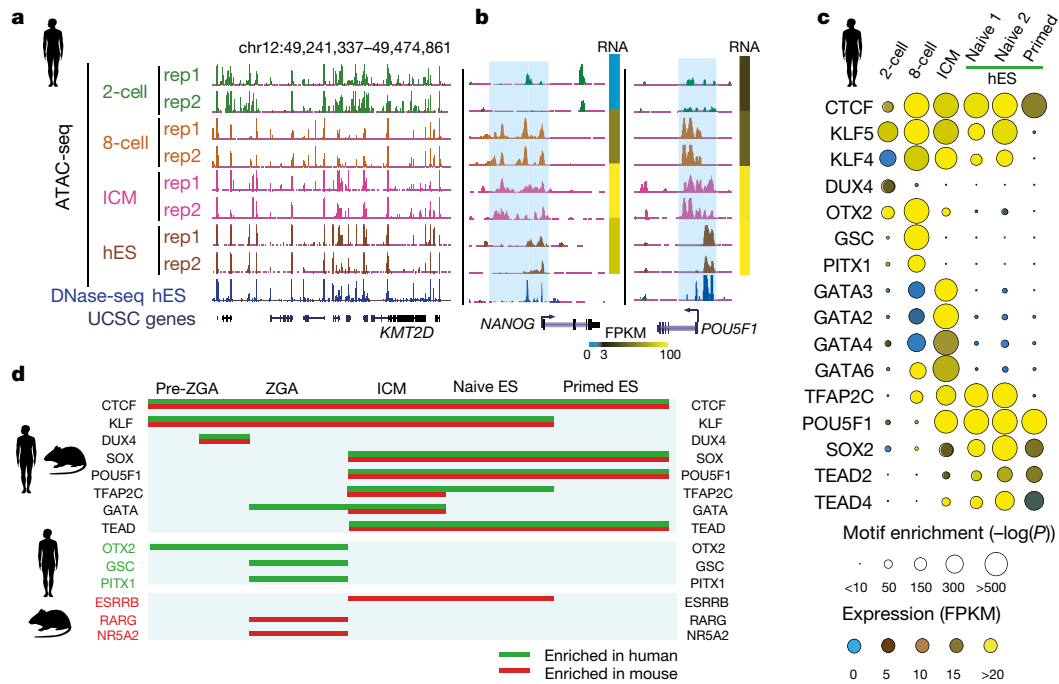


Fig. 1 | Accessible chromatin landscape in human preimplantation embryos. **a**, The UCSC browser view showing ATAC-seq signals in human early embryos (two biological replicates; rep 1 and rep 2). **b**, The UCSC browser views and heat maps showing ATAC-seq enrichment near representative genes and their expression levels detected by RNA-seq. Open chromatin regions are shaded. FPKM, fragments per kilobase of transcript per million mapped reads. **c**, Transcription factor motifs

factors, but do not serve as conventional enhancers to regulate nearby genes (discussed later). Next, HOMER¹¹ analysis unveiled highly stage-specific transcription factor motifs in distal peaks (Fig. 1c). For example, ‘maternal transcription factors’ (two-cell stage and onwards) includes CTCF, KLF and OTX2. ‘Minor ZGA transcription factor’ (two-cell stage) includes DUX4, a key activator of early ZGA genes¹². ‘Major ZGA transcription factor’ (eight-cell and onwards) includes GSC, PITX1, TFAP2C and GATA factors. ‘Pluripotency transcription factors’ (ICMs and ES cells) contains SOX2 and POU5F1. Notably, although conventional ES cells are typically derived from ICMs, they are generally in a primed pluripotent state¹⁰. However, ES cells could also be maintained in a naive state that corresponds to an earlier developmental stage¹⁰. Therefore, we performed ATAC-seq for naive human ES cells^{13,14} (Methods). Interestingly, both KLF and TFAP2C are strongly enriched in ICMs and naive ES cells but not primed ES cells (Fig. 1c). KLF factors are known to have crucial roles in naive pluripotency^{10,15}. TFAP2C is a key germline transcription factor and also a regulator for trophoblast in mice¹⁶. Interestingly, *TFAP2C* is highly expressed in all lineages in human blastocysts (Extended Data Fig. 5c), suggesting human-specific functions. Finally, putative enhancer usages near *POU5F1* in naive ES cells are also more similar to the ICM than primed ES cells (Extended Data Fig. 5d). Notably, both naive and primed ES cells differ from the ICM as they lack the GATA factors that regulate the primitive endoderm lineage¹⁷. These data indicate that the regulatory network of naive human ES cells is indeed more similar to that of ICMs.

We then compared human and mouse transcriptional programs using previous results or ATAC-seq datasets from mouse early embryos¹⁸, mouse ES cells (naive)¹⁸ and mouse epiblast stem cells (EpiSCs; primed)¹⁹. CTCF, KLF, SOX2, POU5F1, GATA and TEAD are all well conserved between human and mouse (Fig. 1d). We also found species-specific transcription factors for preimplantation development including GSC, OTX2 and PITX1 (human) and *ESRRB*, *NR5A2* and *RARG* (mouse)¹⁸ (Fig. 1d). Notably, GSC (mesendoderm), OTX2 (neural lineage) and PITX1 (limb development) are all lineage regulators.

identified from distal ATAC-seq peaks at each stage in human embryos (two-cell, eight-cell and ICM) and human ES cells (primed, naive 1¹⁴ and naive 2¹³). **d**, Schematic representation of putative regulatory transcription factors in early development for those shared by human and mouse as well as those specific to each species. Only enriched and expressed transcription factors are shown.

KLF motifs are again enriched in ICMs and mouse ES cells but not in mouse EpiSCs. Hence, human and mouse have both conserved and divergent regulatory circuitry for early development.

We also observed widespread open chromatin at the two-cell stage (Fig. 1a). To determine whether this may be associated with minor ZGA⁷, we identified genes ($n = 1,018$) with increased transcript levels at the two-cell stage compared to metaphase II (MII) oocytes. Unexpectedly, most of these transcripts are also present in germinal vesicle oocytes in our study and also in MII oocyte using a total RNA-seq dataset²⁰ (Fig. 2a, left and 2b). Importantly, the eight-cell-specific genes, but not the two-cell-specific genes, are sensitive to the transcription inhibitor α -amanitin (Extended Data Fig. 6a and Methods). Consistently, two-cell-specific genes on average only show basal levels of promoter ATAC-seq enrichment (Extended Data Fig. 6b). Shortening poly(A) tails in oocytes is thought to prevent maternal mRNA translation without triggering destabilization²¹. Thus, post-transcription regulation, such as deadenylation in oocytes and polyadenylation in early embryos, may account for a large portion of the differences between mRNA-seq data (this study and earlier studies^{22,23}) and total RNA-seq²⁰. After excluding maternal genes (expressed in germinal vesicle or MII oocytes), we were able to identify 75 potential minor ZGA genes activated at the two-to-four-cell stages (Extended Data Fig. 6c), and this is far fewer than the number of accessible ATAC-seq peaks we detected. However, it is worth noting that we cannot rule out the possibilities that low-input RNA-seq did not fully capture pre-ZGA transcriptional activities. We then asked whether two-cell accessible chromatin may poise genes for activation. Indeed, among all two-cell accessible promoters (maternal genes excluded) ($n = 1,891$), 86% are shared by the eight-cell embryos (Fig. 2c). These promoters are preferentially CpG-rich and transcribed at the eight-cell stage (Fig. 2a, right, and Fig. 2c). Conversely, eight-cell ZGA genes preferentially have accessible promoters at the two-cell stage (44% versus 29% for random genes). Notably, this is not observed for two-cell-specific open promoters (Fig. 2c and Extended Data Fig. 6d). An expanded analysis revealed that these promoters tend to reside in DNA hypomethylated regions in

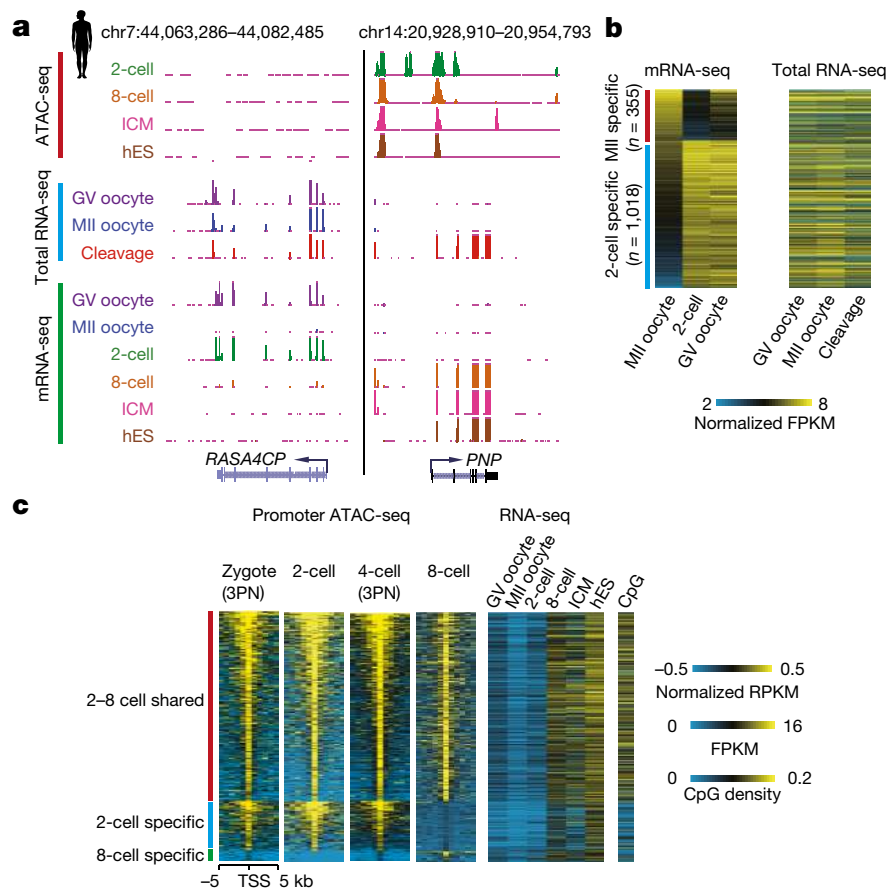


Fig. 2 | Transcription and promoter regulation in human early development. **a**, The UCSC browser view for genes showing early transcripts but without accessible promoters (left), or vice versa (right) at the two-cell stage in humans. Total RNA-seq data were previously published²⁰. **b**, Heat maps showing the expression levels (mRNA-seq) of differentially expressed genes between MII oocytes and two-cell embryos (left). Their expression level in oocyte and early embryos detected by total RNA-seq²⁰ are also shown (right). Cleavage stage denotes pooled two-to-eight-cell

embryos. **c**, Heat maps showing ATAC-seq enrichment around accessible promoters identified at both the two- and eight-cell stages ($n = 1,444$), only at the two-cell stage ($n = 375$) or only at the eight-cell stage ($n = 82$) (left). Gene expression levels and promoter CpG density (number of CpGs per base pair) are also shown (right). Maternal genes (FPKM > 1 in either MII or germinal vesicle (GV) oocytes) were excluded to avoid confounding effects. TSS, transcription start site.

oocytes and blastocyst (despite its global hypomethylation)²⁴ (Extended Data Fig. 6e) (discussed later). Finally, by performing ATAC-seq on 3PN one-cell and four-cell embryos, we found that accessible chromatin at both the one- and four-cell stages highly resembles that of the two-cell stage (Fig. 2c and Extended Data Fig. 7a–c). Hence, pre-ZGA accessible chromatin is established in zygote and transits to a post-ZGA state at the eight-cell stage. In summary, these results suggest that some, but not all, pre-ZGA accessible promoters may poise genes for activation.

Interestingly, two-cell embryos have abundant distal ATAC-seq peaks ($n = 14,770$) with low CpG densities (Fig. 3a and Extended Data Fig. 8a), raising the question of whether they reflect the binding of transcription factors before ZGA⁷. In total 32% of them are shared by the eight-cell embryos and they are indeed enriched for motifs of maternal transcription factors such as CTCF, KLF and OTX2 (Extended Data Fig. 8b). The rest only occur at the two-cell stage and seem unlikely to poise chromatin for local gene activation (Extended Data Figs. 5b and 8b). These loci still enrich for enhancer marks (H3K4me1 and H3K27ac) in human ES cells and transcription factor-binding sites²⁵, albeit at lower levels (Fig. 3a,b and Extended Data Fig. 8c). These peaks tend to reside in DNA hypomethylated regions in human oocyte and blastocyst²⁴ (Fig. 3a, c). Similar patterns are found for the one- and four-cell stages, but are not for ATAC-seq peaks that are specific for other stages or methylomes of sperm or human ES cells. In human oocytes, untranscribed regions are poorly methylated²⁴, forming partially methylated domains (PMDs). We found that 73% of two-cell

peaks are present in human oocyte PMDs (Extended Data Fig. 8d). Next, we asked whether the switch of chromatin accessibility during ZGA depends on transcription. We treated 3PN human zygotes with α -amanitin (with derived embryos term as transcription-blocked embryos, or TBEs). Both TBEs and controls developed normally to eight-cell embryos, which were collected at the late eight-cell stage for ATAC-seq analyses. We found that ATAC-seq enrichment in TBEs is highly similar to pre-ZGA embryos but is distinct from post-ZGA embryos (Fig. 3a, d). Distal open chromatin of TBEs, but not controls, is strongly enriched in oocyte PMDs (Extended Data Fig. 8e, f). Thus, the transition of accessible chromatin landscapes during human ZGA is transcription dependent.

We and others previously showed that in the mouse, oocyte PMDs are uniquely marked by non-canonical forms of histone modifications H3K4me3 and H3K27me3 (but preferentially in non-overlapping sub-regions)^{3,4,26}. Depletion of H3K4me3 results in silencing defects in mature oocytes^{3,5}. Non-canonical H3K4me3 (nH3K4me3) is further inherited to mouse early embryos after fertilization, before being rapidly erased after ZGA possibly by the activated demethylases KDM5A and KDM5B⁴. After ZGA, human *KDM5B* is similarly induced (Extended Data Fig. 9a) and H3K4me3 is decreased based on immunofluorescence analysis^{4,27}. Therefore, we asked whether pre-ZGA accessible chromatin may be associated with histone modifications such as nH3K4me3. As it is challenging to perform chromatin immunoprecipitation followed by sequencing (ChIP-seq) in human embryos, we compared histone modifications in the mouse

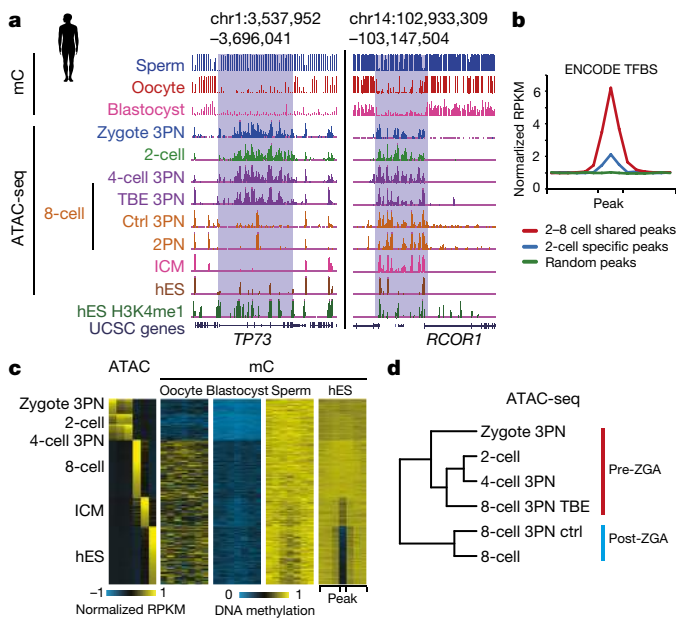


Fig. 3 | Distal accessible chromatin in human early development. **a**, The UCSC browser view showing DNA methylation levels (methylcytosine, or mC) in human sperm, oocyte²⁴, blastocyst²⁴, ATAC-seq enrichment in human early embryos, TBEs and ES cells, and H3K4me1 in human ES cells (ENCODE). Ctrl, control. **b**, The enrichment of ENCODE transcription factor-binding site (TFBS) density around two-cell-specific or two-to-eight-cell shared distal peaks. A random set of peaks that match the lengths of individual two-cell-specific peaks was used as a control. The upstream and downstream regions are $3 \times$ peak lengths away from a peak boundary. **c**, Heat maps showing the oocyte, blastocyst, sperm and ES cell (ENCODE) DNA methylation levels around stage-specific distal ATAC-seq peaks (zygote $n = 889$, two-cell $n = 622$, four-cell $n = 930$, eight-cell $n = 3,382$, ICM $n = 1,777$, ES cell $n = 3,482$). The upstream and downstream regions are $2 \times$ peak lengths away from a peak boundary. **d**, Hierarchical clustering results of whole-genome accessible chromatin states for human early embryos and TBEs.

zygote³ to chromatin accessibility detected by DNase-seq in isolated mouse pronuclei²⁸. Notably, distal open chromatin in maternal pronuclei, but not paternal pronuclei, strongly overlaps with oocyte PMDs and maternal nH3K4me3 in mouse zygotes (Fig. 4a and Extended Data Fig. 9b). Both open chromatin and nH3K4me3 are absent from PMDs after ZGA. Similar correlation could be extended to the paternal allele. We found the paternal H3K4me3, which is re-established after fertilization and briefly exists before ZGA³, also co-localizes with paternal open chromatin (Extended Data Fig. 9c, d). Finally, we found that in full-grown oocytes, DNase-seq peaks also preferentially reside in oocyte PMDs and overlap with nH3K4me3 (Fig. 4a). Thus, these data demonstrate a surprising correlation between non-canonical H3K4me3 and open chromatin in mouse mature oocytes and pre-ZGA embryos.

To determine whether the accessible chromatin transition during mouse ZGA is also transcription-dependent, we treated pronuclear stage 3 (PN3) zygotes with α -amanitin, which arrested mouse embryos at the late two-cell stage. We collected TBEs for ATAC-seq analyses when the control group developed to a later stage (eight-cells, after 45 h). On the basis of single nucleotide polymorphisms (SNPs) between the parental strains, we found the ATAC-seq enrichment in TBEs is similar to that of zygote DNase-seq in an allele-specific manner (Extended Data Fig. 10a–c). Distal open chromatin of TBEs on the maternal allele persisted in oocyte PMDs in which nH3K4me3 is also retained after inhibition of transcription³ (Fig. 4a). The correlation between accessible chromatin and H3K4me3 again extends to the paternal genome in TBEs (Extended Data Fig. 10a, d). Thus, the transition of both accessible chromatin and H3K4me3 during mouse ZGA depends on transcription. Interestingly, despite allele-specific accessibility, we

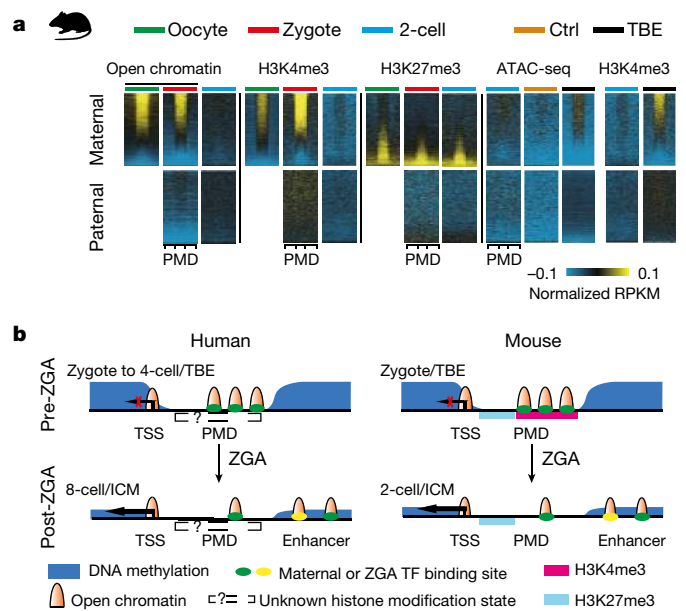


Fig. 4 | Conservation of chromatin transition during ZGA in mouse and human. **a**, Heat maps showing allelic open chromatin (DNase-seq (oocyte and zygote)²⁸ or ATAC-seq (other samples)), H3K4me3³ and H3K27me3²⁶ enrichment around the oocyte PMDs ($n = 24,241$) in mouse oocyte, zygote, 2-cell embryo, TBE and its control. The upstream and downstream regions are $1 \times$ PMD length away from the PMD boundary. **b**, A model shows the transcription and chromatin states before and after ZGA in human and mouse. In pre-ZGA embryos or TBEs, accessible chromatin occurs at CpG-rich promoters and distal regions preferentially in oocyte PMDs. After ZGA, although high-CpG accessible promoters remain open, a large fraction of pre-ZGA distal accessible regions are lost. Newly established accessible chromatin mainly occurs at putative regulatory elements such as enhancers. In mouse, dynamics of distal accessible chromatin also correlates with H3K4me3 in an allele-specific manner (not shown) before and after ZGA. Histone modification states in human are unknown.

identified very similar motifs in distal accessible chromatin on both alleles in zygotes and TBEs, mainly for maternal factors (CTCF and KLFs, such as KLF7, KLF11 and KLF17) (Extended Data Fig. 10e). As a control, transcription factors activated after ZGA (RARG, NR5A2, ESRRB, TEAD4 and GATA)¹⁸ are not enriched, suggesting that similar maternal transcription factors interact with allele-specific chromatin landscapes before ZGA.

Using highly sensitive ATAC-seq, our data revealed both conserved and species-specific transcriptional regulatory networks in human and mouse early embryos. An unusual pre-ZGA chromatin state exists in both human and mouse, featured by extensive accessible chromatin despite the lack of transcription (Fig. 4b). Although accessible promoters preferentially enrich for CpGs, distal accessible chromatin strongly overlaps with oocyte PMDs and, in the case of mouse, nH3K4me3. Transcription is crucial for the transition of H3K4me3 and chromatin accessibility upon ZGA. One critical question is whether pre-ZGA epigenomes in turn regulate transcription or the genome silencing. The accessible regions shared between pre- and post-ZGA stages may have a role in poised activation. By contrast, repressive marks can prevent faulty transcription during drastic epigenetic reprogramming. For instance, oocyte-deposited H3K27me3 is suggested to restrict enhancer function in fly early embryos²⁹ and regulate X-chromosome inactivation and allelic expression in mouse embryos^{28,30}. Interestingly, maternal H3K4me3 also facilitates genome silencing in oocytes^{3–5}. It is possible that nH3K4me3 may work as sponges to facilitate docking or sequestering transcription resources when the genome is silenced. Alternatively, accessible chromatin may be created by unknown factors as ‘chromatin harbours’ that allow the deposition of nH3K4me3 and docking of transcription factors. Future studies are warranted to test

these possibilities. Taken together, our study not only presents a global view for dynamic chromatin landscape in human embryos, but also revealed epigenomic transition during ZGA that are likely to be conserved between human and mouse.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0080-8>.

Received: 27 July 2017; Accepted: 26 March 2018;

Published online: 02 May 2018

- Burton, A. & Torres-Padilla, M. E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 723–735 (2014).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
- Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
- Andreu-Vieyra, C. V. et al. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol.* **8**, <https://doi.org/10.1371/journal.pbio.1000453> (2010).
- Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
- Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).
- Fenouil, R. et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
- Gerdes, P., Richardson, S. R., Mager, D. L. & Faulkner, G. J. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol.* **17**, 100 (2016).
- Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Iturbide, A. & Torres-Padilla, M. E. Starting embryonic transcription for the first time. *Nat. Genet.* **49**, 820–821 (2017).
- Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
- Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
- Guo, G. et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
- Kuckenberger, P., Kubaczka, C. & Schorle, H. The role of transcription factor Tcfap2c/TFAP2C in trophoblast development. *Reprod. Biomed. Online* **25**, 12–20 (2012).
- Morrissey, E. E. et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12**, 3579–3590 (1998).
- Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
- Neijts, R. et al. Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016).
- Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
- Paynton, B. V. & Bachvarova, R. Polyadenylation and deadenylation of maternal mRNAs during oocyte growth and maturation in the mouse. *Mol. Reprod. Dev.* **37**, 172–180 (1994).
- Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
- Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
- Okoe, H. et al. Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* **10**, e1004868 (2014).
- Gerstein, M. B. et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
- Zheng, H. et al. Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol. Cell* **63**, 1066–1079 (2016).
- Zhang, A. et al. Dynamic changes of histone H3 trimethylated at positions K4 and K27 in human oocytes and preimplantation embryos. *Fertil. Steril.* **98**, 1009–1016 (2012).
- Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419–424 (2017).
- Zenk, F. et al. Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212–216 (2017).
- Inoue, A., Jiang, L., Lu, F. & Zhang, Y. Genomic imprinting of *Xist* by maternal H3K27me3. *Genes Dev.* **31**, 1927–1932 (2017).

Acknowledgements We appreciate the laboratory members' comments during preparation of the manuscript. We are grateful for the animal core facility, the sequencing core facility, and bioinformatics facility at Tsinghua University. This work was supported by the National Key R&D Program of China (2016YFC0900300 to W.X. and Y.S., 2017YFA0102802 to J.N.), the National Basic Research Program of China (2015CB856201 to W.X.), the National Natural Science Foundation of China (31422031 and 31725018 to W.X., 31471404 to Y.S., 91740115 to J.N., and 31501205 to J.X.), and the THU-PKU Center for Life Sciences (W.X.). W.X. is a recipient of HHMI International Research Scholar.

Reviewer information Nature thanks I. Hyun and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.S. and W.X. conceived and designed the project. J.W. and B.L. developed miniATAC-seq. B.L., J.W. and J.X. performed the ATAC-seq library construction and sequencing. J.X., W.N. and N.Z. performed RNA-seq library construction and sequencing. J.W. and B.L. analysed the data. G.Y., S.S., S.D. T.L. and X.Z. collected the human oocytes, sperm and 3PN embryos, and performed ICSI and ICM separation. J.X., S.S. and X.Z. performed α -amanitin treatment experiments. Y.G., L.H., W.S., H.J. and L.F. recruited the oocyte and sperm volunteers. P.W., Z.L., B.H. and J.M. performed the mouse embryo experiments. F.D. and X.W. provided primed and naive human ES cells. Q.W. and Y.L. performed NGS sequencing. W.L., J.N., Y.S. and W.X. supervised the project or various experiments. J.W., J.X., B.L. and W.X. wrote the manuscript with the help from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0080-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0080-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.N. or W.X. or Y.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size, and investigators were not blinded to allocation during experiments and outcome assessment.

Ethics statement. This study was approved by the Institutional Review Board (IRB) of The First Affiliated Hospital of Zhengzhou University (2015KY-NO.31) and Tsinghua University (20170009), China. In accordance with the measures of the People's Republic of China on the administration of Human Assisted Reproductive Technology, the ethical principles of the Human Assisted Reproductive Technology and the Human Sperm Bank as well as the Helsinki declaration. The research followed the guiding principles of the Human Embryonic Stem Cell Ethics issued by the MOST and MOH and was regularly reviewed by the Medical Ethics Committee of The First Affiliated Hospital of Zhengzhou University. All gametes and embryos were collected voluntarily after obtaining written informed consent signed by donor couples at the Center for Reproductive Medicine in The First Affiliated Hospital of Zhengzhou University. Oocytes were obtained from the female tubal factors of infertility volunteers who have already had one or two healthy children from natural pregnancy or IVF. The eight-cell and blastocyst stages embryos were donated by couples who already had a healthy baby through IVF, with written informed consents. In the process, couples were informed that their gametes and embryos would be used to study the chromatin state and its regulation during human embryo development, and that their donation would not affect their IVF cycle.

Human gamete and early embryo collection. All gametes and embryos were collected from volunteers who were between 25 and 30 years old, with normal chromosome karyotypes and no family heredity case history, contagion case and smoke history. Healthy sperm were cryopreserved in liquid nitrogen using cryoprotectants. SpermGrad (Vitrolife) was used for gradient sperm separation. For oocytes, the controlled ovarian stimulation were carried out using GnRH analogues combined with human menopausal gonadotrophins or recombinant follicle stimulating hormone (FSH) for pituitary desensitization. Transvaginal ultrasound-guided oocyte collection was scheduled 36 h after hCG administration. Intra cytoplasmic sperm injection (ICSI) was used in this investigation rather than classical IVF to prevent contamination of sperm and cumulus cells. Fertilization was assessed 17–20 h after ICSI and embryo cleavage was recorded every 24 h. Two-cell stages embryos were frozen for storage and sample pooling, before being thawed and subjected to ATAC-seq and RNA-seq. All two-cell and eight-cell embryos and blastocysts used in this research have normal development rates and morphology. The zona pellucida of all embryos used in this study was mechanically removed to avoid blocking the reaction of Tn5 and the potential contamination of cumulus cells. The ICMs of day 5 blastocysts were isolated mechanically according to a method previously reported³¹. In brief, the biopsy pipette was inserted into the blastocoel, and the ICM cells were sucked in and gently pulled out with laser assistance. Numbers of embryos used for this study (2PN) are: two-cell rep1: 15; two-cell rep2: 17; eight-cell rep1: 8; eight-cell rep2: 10; ICM rep1: 3; ICM rep2: 3. For the 3PN embryos, the 3PN zygotes were identified after ICSI or IVF, which were collected at the one-cell stage or further cultured to the two-, four- or eight-cell stage and blastocyst using G-1 (Vitrolife) human embryos culture medium. G-2 (Vitrolife) medium was used from the eight-cell to blastocyst stage as normal 2PN embryos. Only the high-quality 3PN embryos and blastocysts without development arrest were collected.

Treatment of human embryos with α -amanitin. The 3PN zygotes were identified and collected after IVF or ICSI. About 50 3PN zygotes were thawed and transferred to G1 media in the presence or absence of α -amanitin (25 ng μ l⁻¹, Sigma-Aldrich). After the control and α -amanitin-treated embryos reaching the eight-cell stage, high-quality embryos were collected and those with fragments or arrested at other stages were discarded. The zona pellucida was gently removed by laser. The cells were then washed three times with PBS to avoid any potential contamination then further subjected to ATAC-seq library preparation.

Mouse embryo collection. All animal experiments were approved by Institutional Animal Care and Use Committee (IACUC) of Tsinghua University, Beijing, China. Pre-implantation embryos were collected from 5–6-week-old C57BL/6N females (Vital River) mated with PWK/Phj males (Jackson Laboratory). The embryo collection was performed as described previously¹⁸. To inhibit transcription in early embryos, PN3 zygotes were cultured in CZB supplemented in the absence or presence of α -amanitin (125 μ g ml⁻¹) for about 45 h. The resulting embryos were subjected to ATAC-seq analyses.

Cell culture of human ES cells. For primed human ES cell culturing, ES cells (H1) were maintained on Matrigel (BD Biosciences)-coated plates (Corning) in E8 medium (STEMCELL Technologies) in a feeder-free and serum-free manner. Naive-state human ES cells were generated as described previously with some modifications^{13,14}. In brief, for t2iL/Gö-human ES cell derivation¹⁴, the H7 human ES cells were seeded on matrigel (BD Biosciences) coated plate 48 h before virus infection. Subsequently, cell culture medium was changed to fresh human ES cell culture medium supplemented with 3 virus concentrates containing FUW-human

NANOG, FUW-human KLF2 and FUW-M2rtTA in the presence of 8 μ g ml⁻¹ polybrene for 24 h. Two days later the human ES cells were passaged using Accutase (Thermo Fisher) and seeded on mouse embryonic fibroblast (MEF) feeder layers in human ES cell culture medium with 10 μ M Y27632. After 24 h, medium was changed to t2iL/Gö medium for further culture. After 7–14 days of culture, colonies with naive mouse ES cell-like morphology appeared and were picked to derive the t2iL/Gö-human ES cells with t2iL/Gö medium. Note that the t2iL/Gö-human ES cell proliferation rate at the initiation stage of the conversion was very low, and initial passage of t2iL/Gö-human ES cells were pursued every 6–7 days using Accutase. The 5iL/A-human ES cell derivation¹³ was similar to the t2iL/Gö-human ES cell derivation, but after virus transfection, the transfected ES cells were directly cultured in 5iL/A medium supplemented with 0.5% KOSR, 8 ng ml⁻¹ FGF2 and 2 μ g ml⁻¹ DOX. After 2 weeks of culture, colonies with naive mouse ES cell-like morphology could be isolated and expanded in 5iL/A medium without KOSR, FGF2 and DOX. All cell culture was performed under conditions of 5% oxygen at 37 °C.

The conventional H7 human ES cells were cultured on mitomycin C-inactivated MEF feeder layers in human ES cell culture medium containing 20% knockout serum replacement (KOSR) (Thermo Fisher), 78% DMEM/F12 medium (Thermo Fisher), 1% NEAA (Thermo Fisher), 2 mM GlutaMax (Thermo Fisher), 0.1 mM β -mercaptoethanol (Thermo Fisher), and 8 ng ml⁻¹ FGF2 (R&D). The t2iL/Gö medium contains DMEM/DF12 and Neurobasal medium mixed at 1:1 ratio, 0.5 \times N2 supplement, 0.5 \times B27 supplement, 2 mM GlutaMax, 1% NEAA, 0.1 mM β -mercaptoethanol, 50 μ g ml⁻¹ BSA, 10 μ g ml⁻¹ insulin, 1 μ M PD0325901, 1 μ M CHIR99021, 50 μ g ml⁻¹ L-ascorbic acid (Sigma-Aldrich), 10 ng ml⁻¹ human LIF, 2.5 μ M Gö6983 (Sigma-Aldrich) and 10 μ M Y27632 (Selleck). The 5iL/A medium contains DMEM/DF12 and Neurobasal medium mixed at 1:1 ratio, 0.5 \times N2 supplement, 0.5 \times B27 supplement, 1 mM GlutaMax, 1% NEAA, 0.1 mM β -mercaptoethanol, 50 μ g ml⁻¹ BSA, 1 μ M PD0325901, 20 ng ml⁻¹ human LIF, 1 μ M IM-12 (Enzo), 0.5 μ M SB590885 (R&D), 1 μ M WH-4-023 (A Chemtek), 10 μ M Y27632 and 20 ng ml⁻¹ activin A (Peprotech).

Plasmid cloning and viral preparation. For construction of lentiviral vectors containing the human NANOG and KLF2 genes, the human NANOG and KLF2 coding sequences were cloned from human ES and fibroblast cell cDNA libraries, and inserted after the TetO/CMV promoter of the FUW lentivirus vector for DOX-inducible expression, respectively. Lentiviral particles containing FUW-human NANOG, FUW-human KLF2 and FUW-M2rtTA (Addgene ID: 20342) were packaged in 293FT cells with a VSVG coat. Viral supernatants were collected at 48 h after transfection, and then were filtered using a 0.45- μ m filter and concentrated by high-speed centrifugation.

miniATAC-seq library generation and sequencing. The miniATAC-seq libraries were prepared as previously described with further modifications primarily for DNA purification^{2,18}. In brief, samples were transferred into 6 μ l lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂ and 0.5% NP-40) for 10 min on ice. After lysis, 4 μ l ddH₂O, 4 μ l 5 \times TTBL, 5 μ l TTE mix V5 (TD502, Vazyme) were added to the sample, which was mixed and incubated at 37 °C for 30 min. The reaction was terminated by incubating with 5 μ l 5 \times TS stop buffer (TD502, Vazyme) at the room temperature for 5 min, before 2 μ l carrier RNA (20 ng μ l⁻¹ after 50 \times dilution) (EpiTect Fast DNA Bisulphite Kit 59824) and 103 μ l Tris-EDTA (TE) buffer was added to the sample. The sample was then transferred to a 1.5-ml tube. After adding 130 μ l phenol–chloroform (vortexed and incubated at room temperature for 3 min), the sample was transferred to a phase-lock tube (WM5-2302820 TIANGEN). After spinning at maximum speed for 15 min, the supernatant was transferred to a new 1.5-ml tube, and 650 μ l ethanol, 24 μ l sodium acetate (3 M) and 2 μ l glycogen were added for DNA precipitation at –20 °C overnight. The next day, DNA pellet was spun down at maximum speed for 15 min at 4 °C, washed with 75% ethanol, air dried, and resuspended in 29 μ l H₂O. The phenol–chloroform purification with carrier RNA improves the efficiency for both DNA purification and removal of residual Tn5 transposase, which may otherwise interfere with subsequent reactions. DNA was transferred to a 0.2-ml PCR tube, and 5 μ l N5XX primer, 5 μ l N7XX primer (TD202, Vazyme), 10 μ l 5 \times TAB and 1 μ l TAE (TD502, Vazyme) were added. PCR was performed to amplify the library for 18 cycles for about 30 embryos or ES cells using the following PCR conditions: 72 °C for 3 min; 98 °C for 30 s; and thermocycling at 98 °C for 15 s, 60 °C for 30 s and 72 °C for 3 min; following by 72 °C 5 min. After the PCR reaction, libraries were purified with the 1.5 \times AMPure (Beckman) beads and were subjected to next-generation sequencing.

RNA-seq library generation and sequencing. The Smart-seq2 libraries of human oocytes and early embryos were prepared using SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Clontech, 634888) also as previously described³². Oocytes and early embryos from the same batch were divided into two parts, with 95% cells used for ATAC-seq and the rest for RNA-seq. Oocytes and early embryos (two-cell, four-cell and eight-cell stage, and ICMs) were lysed in lysis buffer containing RNase inhibitor according to the user manual.

The library was quantified using Qubit and Agilent 2100 before being subjected to sequencing.

ATAC-seq data processing. The single-end ATAC-seq reads were aligned to mm9 or hg19 reference genome with random chromosome cleaned by Bowtie (version 2.2.2)³³ under the parameters `-t -q -N 1 -L 25`. The paired-end ATAC-seq reads were aligned with the parameters: `-t -q -N 1 -L 25 -X 2000 -no-mixed -no-discordant`. All unmapped reads, non-uniquely mapped reads and PCR duplicates were removed. For downstream analysis, we normalized the read counts by computing the numbers of reads per kilobase of bin per million of reads sequenced (RPKM). RPKM values were averaged for each bin between replicates. To minimize the batch and cell type variation, the RPKM values were further normalized by Z-score transformation. To visualize the ATAC-seq signal in the UCSC genome browser, we extended each read by 250 base pairs (bp) and counted the coverage for each base. The correlation between ATAC-seq replicates was calculated as following: each read was extended 250 bp from the mapped end position and the RPKM value was generated on a 100 bp-window base. The ATAC-seq enrichment was then summed within each 2-kb window for the entire genome and was compared between replicates. Pearson correlation was calculated and was shown. Allele assignment of sequencing reads for mouse TBE embryos was conducted as described previously¹⁸. In brief, to assign each read to its parental origins, we examined all SNPs in the read that showed high-quality base calling (Phred score ≥ 30). For paired-end reads, SNP information from both reads in the pair was summed and used. When multiple SNPs were present in a read (or a read pair), the parental origin was determined by votes from all SNPs and the read was assigned to the allele that had at least two thirds of the total votes.

Gene expression data processing. All RNA-seq data were mapped to hg19 genome by Tophat (version 2.4.0)³⁴. The gene expression level was calculated by Cufflinks (version 2.2.1)³⁵ based on the hg19 refFlat annotation database from the UCSC genome browser. The Spearman correlation coefficients between RNA-seq replicates were calculated. For published total RNA-seq data²⁰, the expression values were quantile normalized together with the mRNA-seq data in this study to allow further comparison. For naive human ES cell gene expression microarray data¹³, the expression values were downloaded and averaged for each gene. The averaged values were quantile normalized together with the FPKM value in this study to allow further comparison.

DNase-seq, ChIP-seq and DNA methylation data processing. Downloaded DNase-seq and ChIP-seq data were mapped to mm9 genome by Bowtie with similar parameters as ATAC-seq data. All unmapped reads, non-uniquely mapped reads and PCR duplicates were removed. For downstream analysis, we calculated the read counts by computing RPKM on genome 100-bp bin. Downloaded DNA methylation data were mapped to mm9 or hg19 genome by BSMAP³⁶ with parameters: `-r 0 -p 16 -w 100 -v 0.1`. PCR duplicates were removed. For each CpG site, the methylation level is calculated as the total methylated counts (combining Watson and Crick strands) divided by the total counts across all reads covering this CpG.

Identification of stage-specific genes. The minor ZGA genes activated at the two-to-four-cell stages were identified by selecting those expressed at the two-cell stage (FPKM > 5) but not expressed or expressed at low levels in MII and germinal vesicle oocytes (FPKM < 5) in both mRNA-seq and total RNA-seq, with a minimal twofold gene expression change between the two- or four-cell and oocyte stages.

For stage-specific genes in eight-cell embryos, ICMs and human ES cells, a Shannon entropy-based method was used to identify stage-specific genes, as previously described³⁷. Owing to the possible confounding effects from maternally inherited RNA transcripts, ZGA-only genes were analysed, which were defined as those not expressed in MII or germinal vesicle oocytes (FPKM < 1). Genes with entropy scores of less than 2 were selected as candidates for stage-specific genes. Among these genes, we selected candidates of stage-specific genes for each stage based on the following criteria: the gene is highly expressed at this stage (FPKM > 10), and such high expression cannot be observed in other stages. These genes were then reported in the final stage-specific gene lists.

Identification of ATAC-seq peaks and their genome coverages. All the ATAC-seq peaks were called by MACS v1.4³⁸ with the parameters `-nolambda -nomodel`. ATAC-seq peaks that are at least 2.5 kb away from annotated promoters from refFlat were selected as distal ATAC-seq peaks. The genome coverages of peaks from different samples were calculated by genomeCoverageBed³⁹ using hg19 reference genome.

The comparison between ATAC-seq peaks and known cis-regulatory elements. To compare the ATAC-seq peaks identified in early embryos with the annotated cis-regulatory elements, we calculated the overlap between the ATAC-seq peaks of different stages and annotated promoters (TSS ± 0.5 kb). Distal peaks were then compared to distal DNase I hypersensitive sites in ES cells. Random peaks were generated by selecting random regions in the genome with the sizes matching each individual ATAC-seq peak.

The comparison between ATAC-seq peaks and repetitive elements. To identify the overlap between repetitive elements and promoter or distal ATAC-seq

peaks, the ATAC-seq peaks were compared with the locations of annotated repeats (RepeatMasker) downloaded from the UCSC genome browser by intersectBed³⁹ with default parameters. As repeats of different classes vary greatly in numbers, a random set of peaks with identical lengths of ATAC-seq peaks was used for the same analysis as a control. The numbers of observed peaks that overlap with repeats were compared to the numbers of random peaks that overlap with repeats, and a log ratio value (\log_2) was generated as the 'observed/expected' enrichment.

Identification of stage-specific distal ATAC-seq peaks. The two-cell-specific distal ATAC-seq peaks were identified as those with high levels of ATAC-seq enrichment at the two-cell stage (normalized RPKM > 1) but not at the eight-cell stage (normalized RPKM < 0). Those with normalized RPKM > 1 at both stages were identified as shared distal peaks between the two- and eight-cell stages.

To identify stage-specific distal ATAC-seq peaks, the distal ATAC-seq peaks from human embryos of all stages and ES cells were combined, with overlapped peaks merged. The average RPKM values were calculated for these distal ATAC-seq peaks which were further normalized by the Z-score normalization. A Shannon entropy-based method³⁷ was used to identify stage-specific distal ATAC-seq peaks. We selected those with entropy less than 2 as candidates for stage-specific distal ATAC-seq peaks. The stage-specific distal ATAC-seq peaks were further defined based on the following criteria: the distal ATAC-seq peak has high enrichment at this stage (normalized RPKM > 1), and positive enrichment (normalized RPKM > 0) at no more than two additional stages. The resulting distal ATAC-seq peaks were then reported in the final stage-specific distal ATAC-seq peak list. The functional enrichment for genes that are near stage specific distal ATAC-seq peaks was analysed using the GREAT tool⁴⁰ by default settings.

Prediction of promoter targets of putative enhancers in distal ATAC-seq peaks. To identify the potential targeted genes for stage specific enhancer (distal peaks), we computed the averaged ATAC-seq enrichment (normalized RPKM) for all distal ATAC-seq peaks and annotated promoters (TSS ± 0.5 kb). Among genes assigned to enhancers by GREAT analysis based on distances, we further calculated the correlation between the ATAC-seq enrichment at distal ATAC peaks and these promoters across human embryos and ES cells. The promoter with a Pearson correlation coefficient above 0.8 was selected as the potential target of the enhancer.

Gene ontology analysis. The DAVID web-tool was used to identify the Gene Ontology (GO) terms using databases including Molecular Functions, Biological Functions and Cellular Components⁴¹.

Hierarchical clustering analysis. The hierarchical clustering was performed in R by hclust() function with ATAC-seq RPKM values via Spearman correlation coefficients.

Motif, enhancer mark and transcription factor-binding sequence analyses for distal ATAC-seq peaks. To find the sequence motifs enriched in distal ATAC-seq peaks, findMotifsGenome.pl from the HOMER program¹¹ was used. Motifs with known match in HOMER database were selected. The ChIP-seq data for human ES cell H3K27ac and H3K4me1 marks⁴² and the collection of transcription factor-binding sites²⁵ were downloaded from the UCSC genome browser. The average RPKM values at the distal peaks and their nearby regions were calculated. The number of transcription factor-binding sites was first binned for each 100-bp window, and the average enrichment at the distal peaks and their nearby regions was calculated.

Identification of oocyte PMDs. PMDs in oocytes were identified as described previously³. In brief, published human²⁴ and mouse⁴³ oocyte DNA methylation datasets were downloaded. Average oocyte DNA methylation levels were calculated for 10-kb bins of the genome. Bins with average DNA methylation levels of lower than 0.5 were selected and merged into PMDs. Promoter regions (± 2.5 kb) were excluded from PMDs.

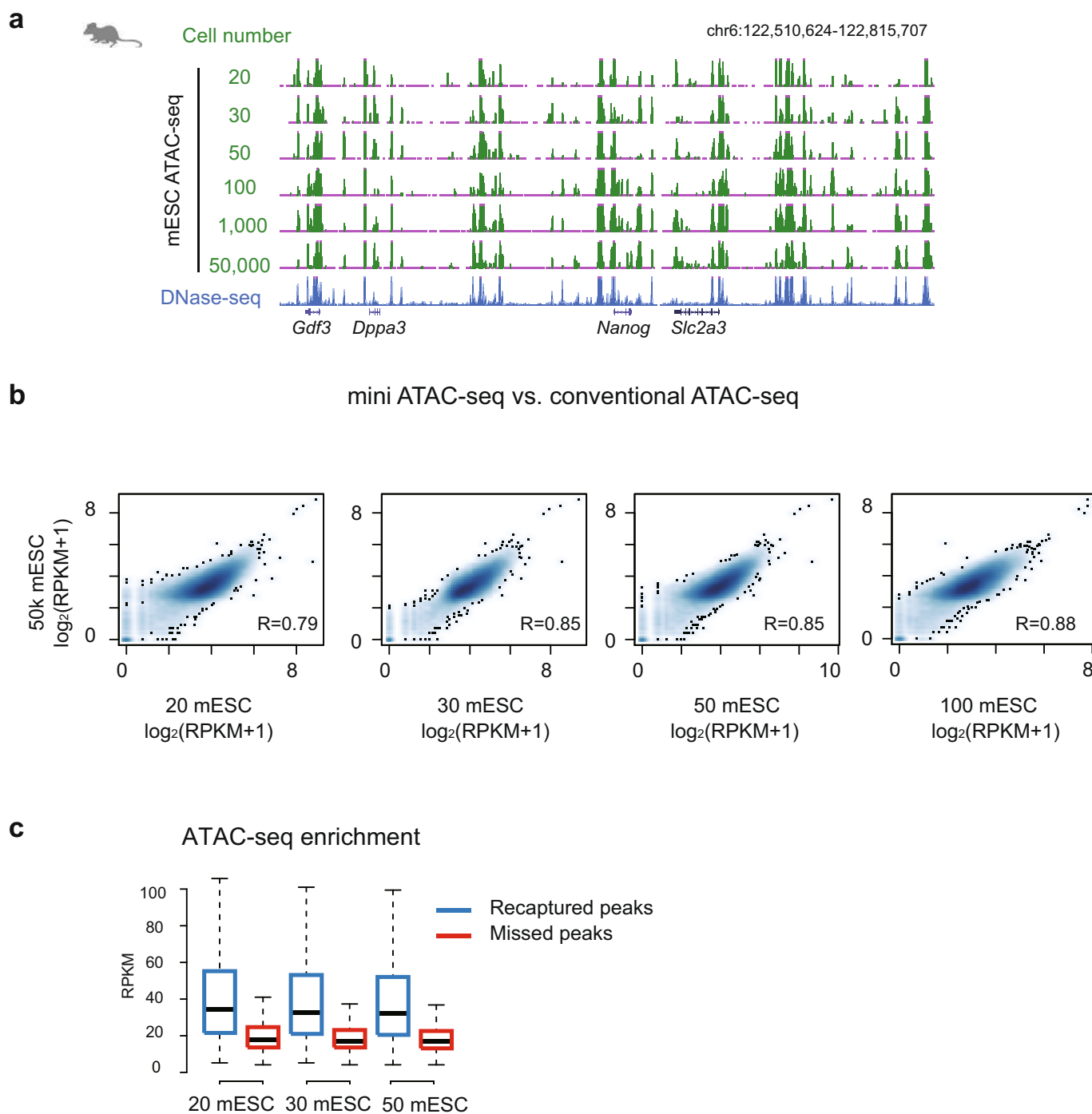
The ATAC-seq enrichment in PMD. ATAC-seq RPKM is Z-score normalized on the genome-wide 100-bp bin with promoter regions removed. PMDs, as well as their upstream and downstream regions, were divided into five bins separately, and the mean RPKM was calculated for each bin and shown in the heat map. The average ATAC-seq RPKM for all PMDs, as well as their upstream and downstream regions, was computed and shown.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data sets generated and analysed during the current study are available in the Gene Expression Omnibus (GEO) with the accession number GSE101571.

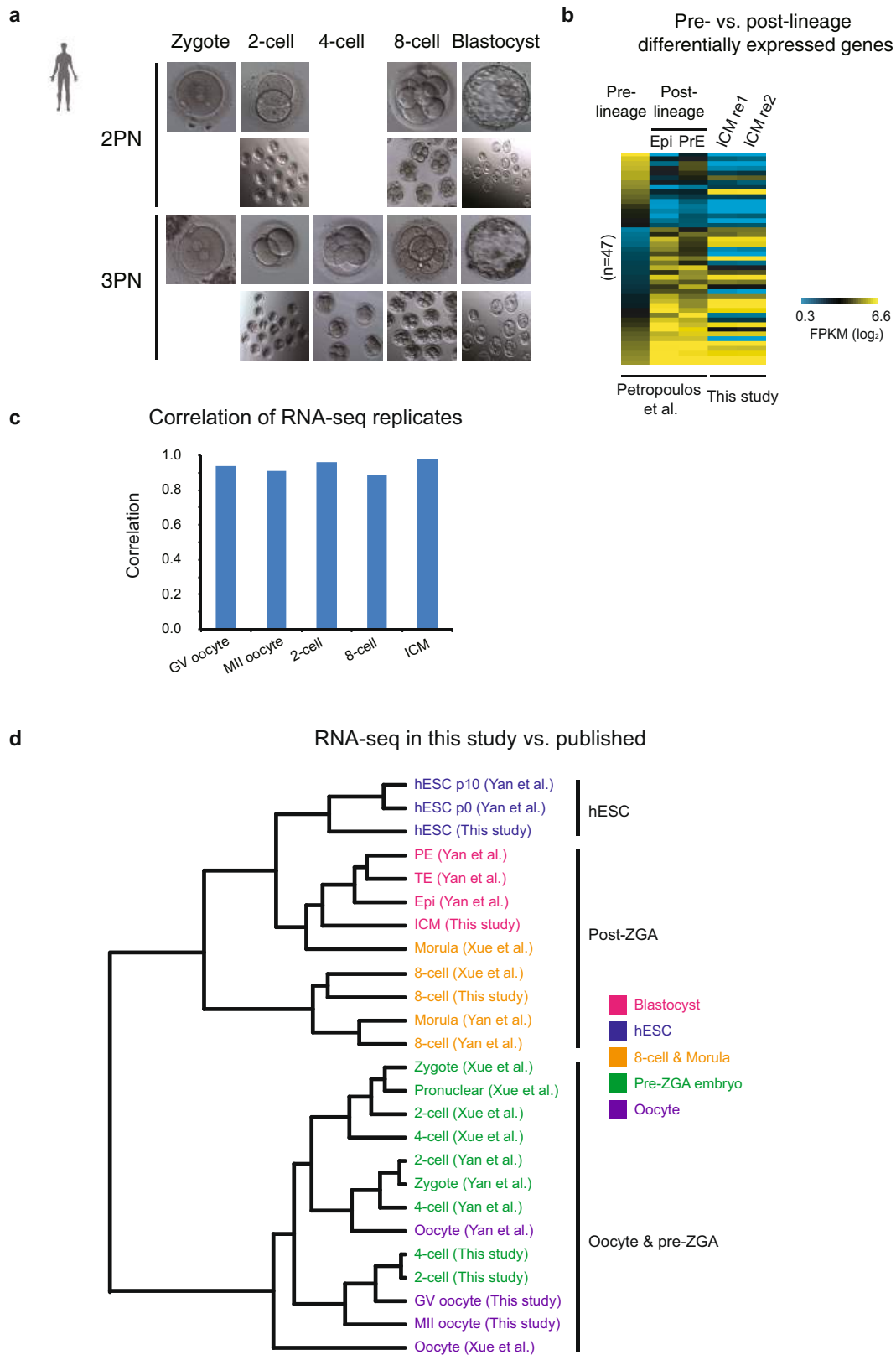
- Capalbo, A. et al. FISH reanalysis of inner cell mass and trophectoderm samples of previously array-CGH screened blastocysts shows high accuracy of diagnosis and no major diagnostic impact of mosaicism at the blastocyst stage. *Hum. Reprod.* **28**, 2298–2307 (2013).
- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

34. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
35. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562–578 (2012).
36. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
37. Schug, J. et al. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).
38. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
40. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
41. Dennis, G., Jr et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
42. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
43. Kobayashi, H. et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8**, e1002440 (2012).
44. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **167**, 285 (2016).
45. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
46. Hawkins, R. D. et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
47. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).



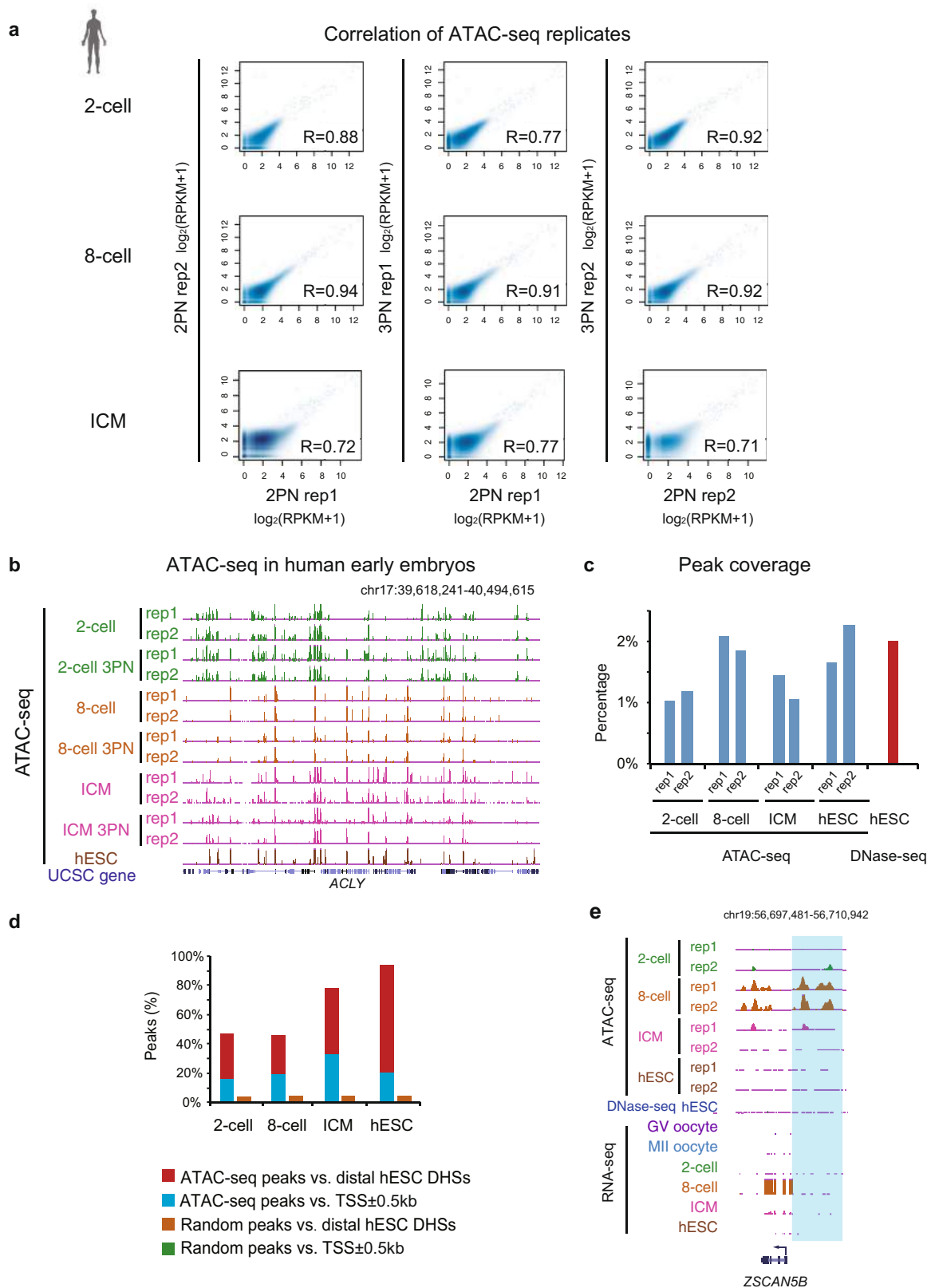
Extended Data Fig. 1 | Validation of miniATAC-seq. **a**, UCSC browser view showing the ATAC-seq signals using various numbers of mouse ES cells. **b**, Scatter plots comparing the miniATAC-seq signals from various numbers of mouse ES cells with conventional ATAC-seq using 50,000

mouse ES cells. The Pearson correlation coefficients are also shown. **c**, Box plots showing the ATAC-seq enrichment for peaks from 50,000 mouse ES cells that are recaptured or missed by miniATAC-seq using lower numbers of mouse ES cells.



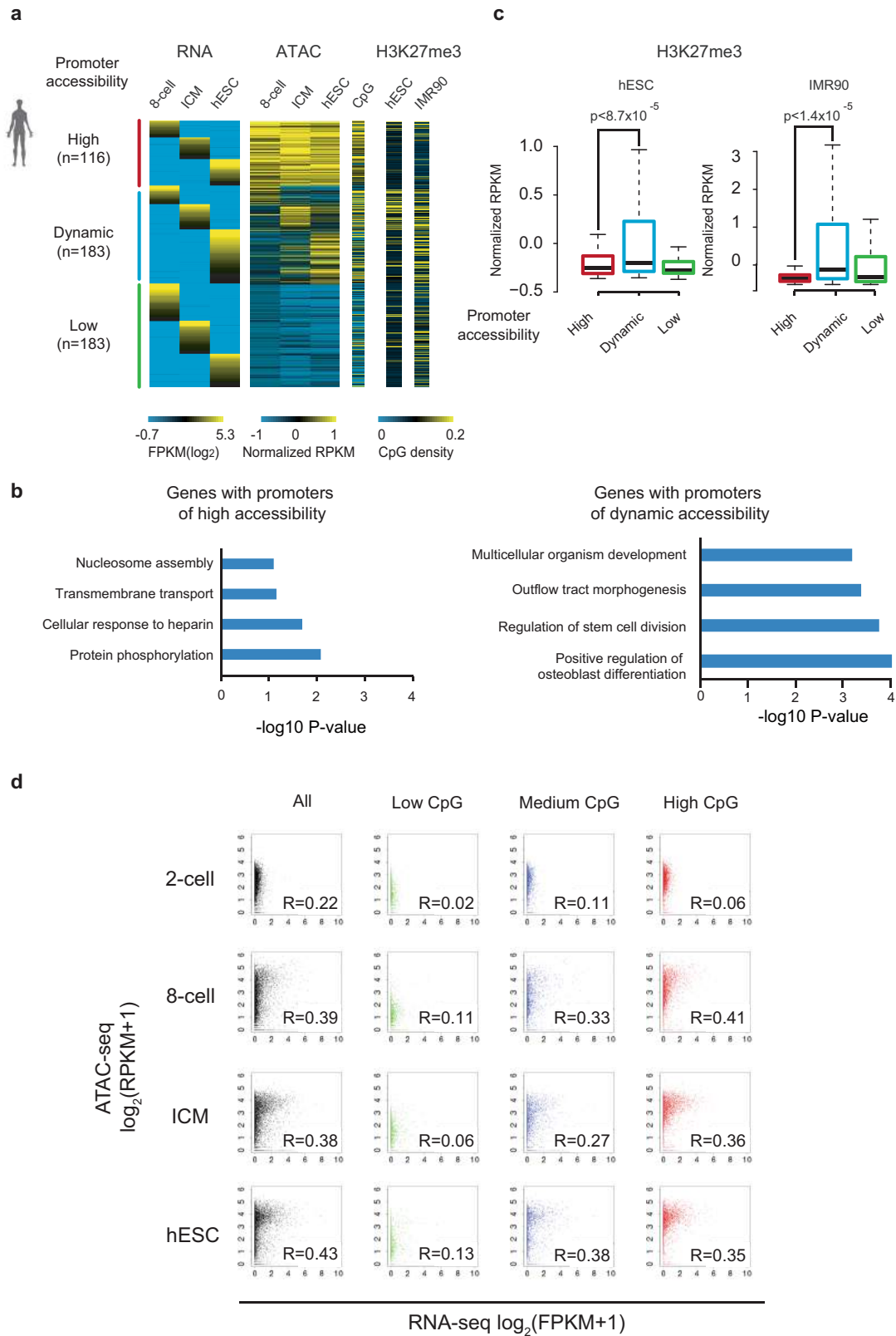
Extended Data Fig. 2 | Validation of RNA-seq data in human early embryos. **a**, Microscopy images of human 2PN and 3PN embryos at the zygote, two-cell, four-cell, eight-cell and blastocyst stages. ICSI was used to avoid the cumulus cell contamination. High-quality score embryos were selected for subsequent study (represented by magnified images). **b**, Heat maps showing gene expression levels for differentially expressed genes between human embryos of pre-lineage segregation and post-lineage

specification⁴⁴. Expression of ICMs in this study is also shown for two replicates. Epi, epiblast, PrE, primitive endoderm. **c**, Bar charts showing the Spearman correlation between the two replicates of RNA-seq samples. **d**, Hierarchical clustering of RNA-seq datasets from this study (using Smart-seq) and previous studies^{22,23} (using a different mRNA-seq method⁴⁵). Pearson correlation was used to measure distances. Different colours represent various stages.



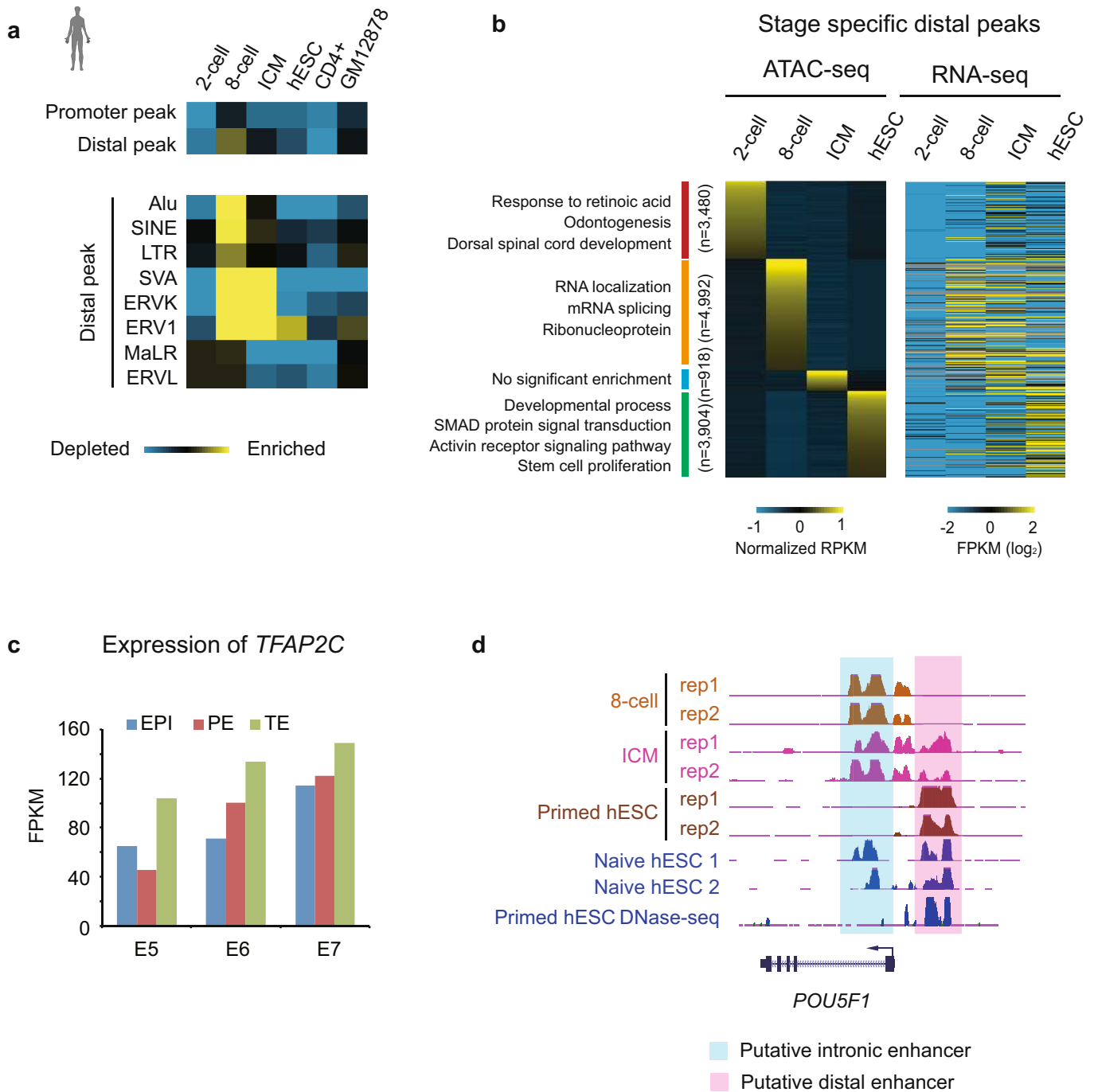
Extended Data Fig. 3 | Validation of ATAC-seq data in human early embryos. **a**, Scatter plots showing the ATAC-seq signals between replicates at each stage in human early development or between 2PN and 3PN embryos. **b**, UCSC browser view showing the landscape of accessible chromatin in replicates of human early embryos. **c**, Bar charts showing the genome coverages of ATAC-seq peaks of each stage or DNase-seq peaks

for human ES cells (ENCODE). **d**, The overlap between ATAC-seq peaks and annotated promoters (TSS ± 0.5 kb) or distal DNase I hypersensitive sites in ES cells. A random set of peaks that match the lengths of individual ATAC-seq peaks were used as a control. **e**, UCSC browser views showing the ATAC-seq and RNA-seq enrichment near a representative gene. Open chromatin regions are shaded.



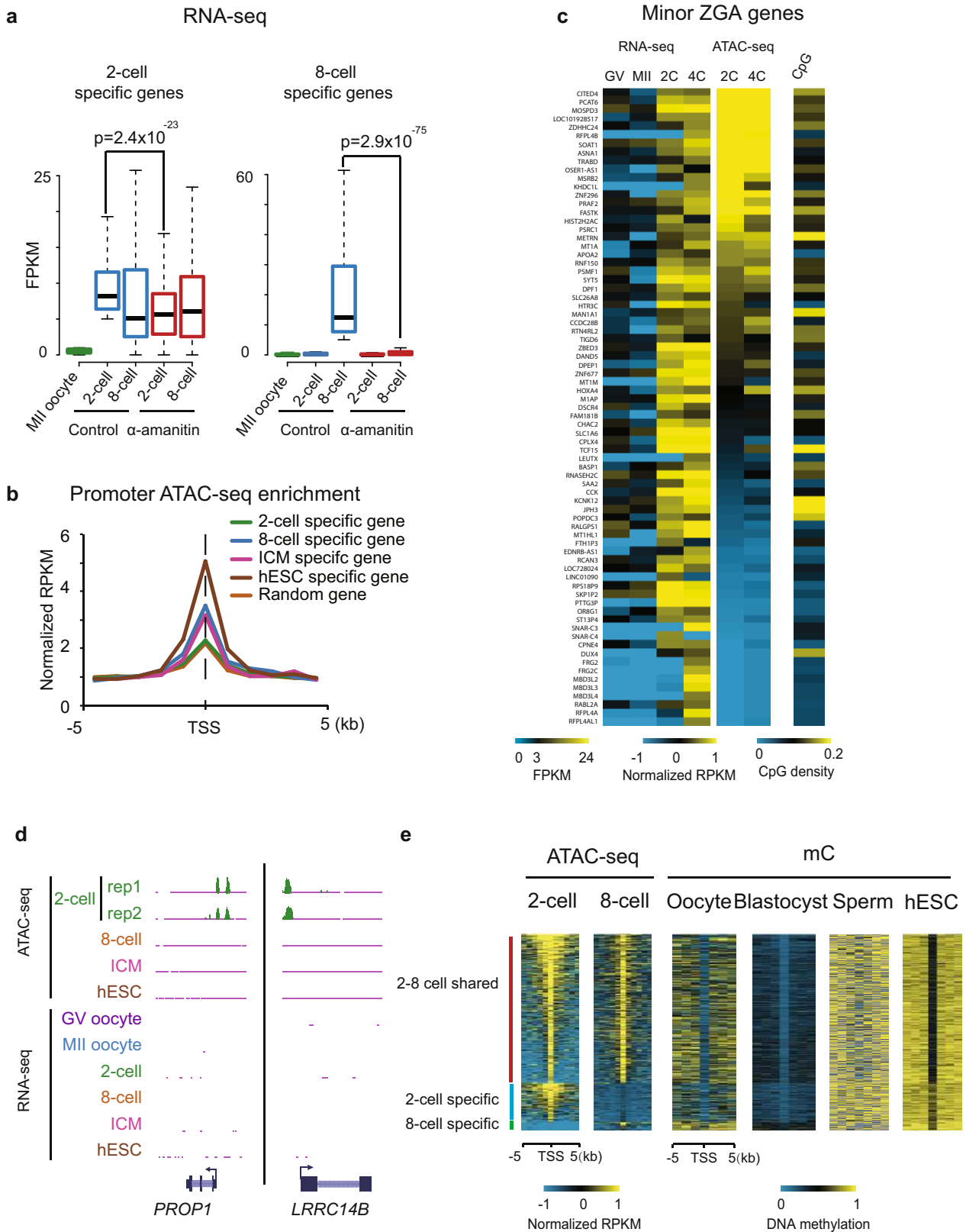
Extended Data Fig. 4 | Relationship of chromatin accessibility and transcription in human early embryos. **a**, Heat maps showing three classes of promoter accessibility (high, dynamic and low) for stage-specific genes (maternal genes excluded). CpG densities and H3K27me3 levels in human ES cells and fibroblasts (IMR90)⁴⁶ are also shown. **b**, GO analysis results for gene classes in **a**. (The 'low' class does not have gene

set enrichment with $P < 10^{-2}$). **c**, Box plots showing promoter enrichment of H3K27me3 in human ES and IMR90 cells⁴⁶ for each class in **a**. P values based on a one-sided t -test are shown. **d**, Scatter plots showing promoter ATAC-seq enrichment and gene expression (maternal genes excluded) for all genes (black) or genes with promoters of low (green), medium (blue), or high (red) CpG densities. Spearman correlation coefficients are shown.



Extended Data Fig. 5 | Features of distal ATAC-seq peaks in human embryos. **a**, Top, enrichment of repeats in ATAC-seq promoter and distal peaks compared to that in random peaks for early human embryos, human ES cells, and somatic cell types. The enrichment was calculated as a \log_2 ratio for the numbers of observed peaks that overlap with repeats divided by the numbers for random peaks. A random set of peaks that match the lengths of individual ATAC-seq peaks was used. Bottom, a similar analysis was performed for the enrichment of repeat subfamilies in distal peaks. **b**, Middle, heat maps showing the ATAC-seq signals for stage-specific

distal ATAC-seq peaks in human embryos. Left, GREAT analysis⁴⁰ results are also shown. Right, the expression of predicted targets among GREAT listed nearby genes for putative enhancers (distal peaks) (Methods) are shown for each stage. **c**, Bar charts showing the expression of *TFAP2C* in epiblast (EPI), primitive endoderm (PE) and trophoctoderm (TE) lineages in human embryos from embryonic days 5–7 based on a previous study⁴⁴. **d**, UCSC browser view showing the ATAC-seq signals around *POU5F1* in human early embryos and ES cells (primed, naive 1¹⁴ and naive 2¹³). Human ES cell DNase-seq data (ENCODE) are shown as a control.



Extended Data Fig. 6 | Transcription and chromatin states before major human ZGA. **a**, Box plot showing the expression levels of two-cell specific (left) and eight-cell specific (right) genes in MII oocytes and embryos with or without α -amanitin treatment. P values based on a one-sided t -test are shown. **b**, The average ATAC-seq enrichment for each stage is shown at the promoters of stage specific genes at the same stage identified by mRNA-seq. Two-cell ATAC-seq enrichment for a random set of promoters were similarly analysed as a control. **c**, Heat maps showing expression levels of possible minor ZGA genes activated at the two-to-four-cell stage for their

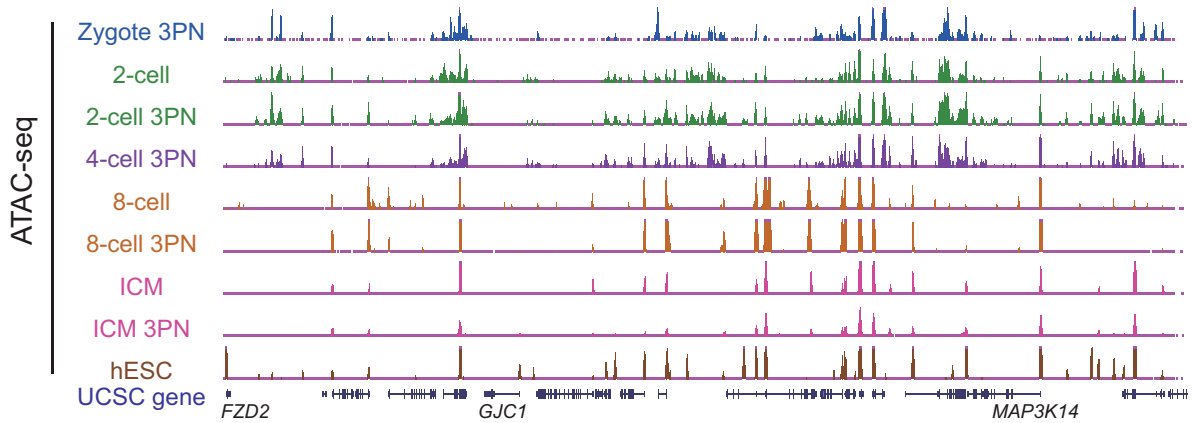
expression in germinal vesicle oocytes, MII oocytes, and two-to-four-cell embryos. Promoter ATAC-seq enrichment for two-to-four-cell stages and CpG densities are also shown. **d**, UCSC browser view showing the promoter ATAC-seq signals specifically appearing in two-cell embryos. **e**, Left, heat maps showing ATAC-seq enrichment at the accessible promoters present in both two- and eight-cell embryos, as well as those specific to each stage. Right, the human oocyte, blastocyst, sperm and ES cell DNA methylation levels around these promoters are also shown.

a



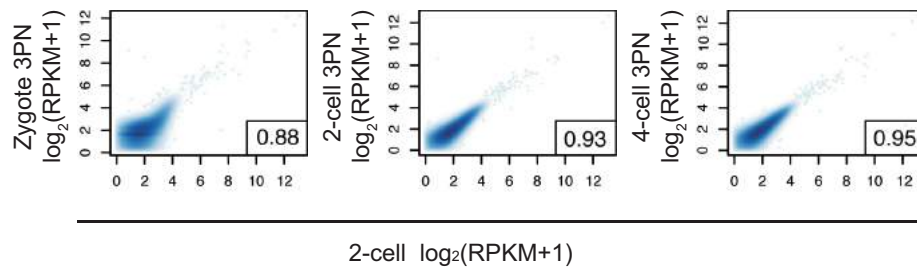
ATAC-seq in normal and 3PN embryos

chr17:42,630,964-43,532,364



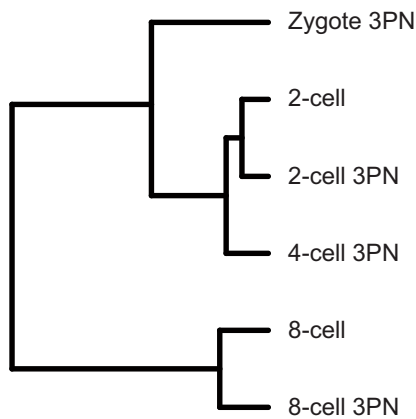
b

ATAC-seq in 1-4 cell embryos



c

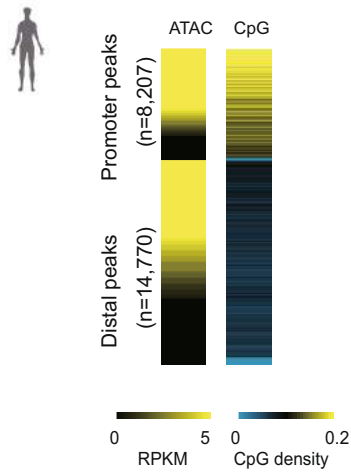
Hierarchical clustering of ATAC-seq signals



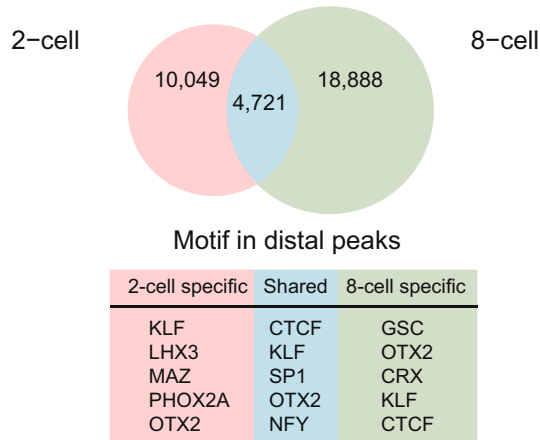
Extended Data Fig. 7 | Accessible chromatin state in one-cell and four-cell human embryos. a, UCSC browser view showing the ATAC-seq signals in human early embryos. b, Scatter plots showing the ATAC-seq signals between two-cell human embryos and 3PN one-cell, two-cell

and four-cell embryos. c, Hierarchical clustering results showing the relationships of chromatin states among human embryos based on whole-genome ATAC-seq enrichment.

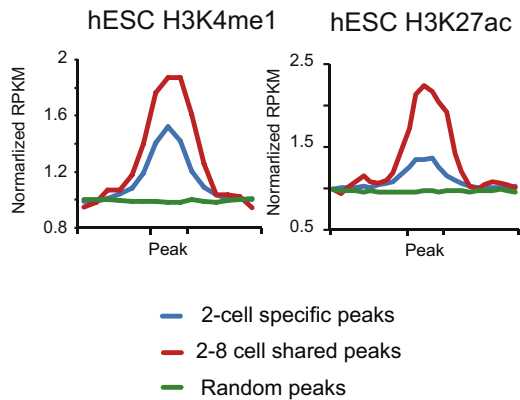
a 2-cell ATAC-seq vs. CpG density



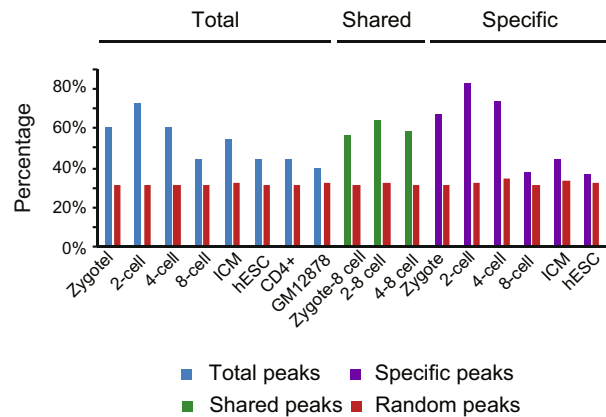
b Distal ATAC-seq peaks



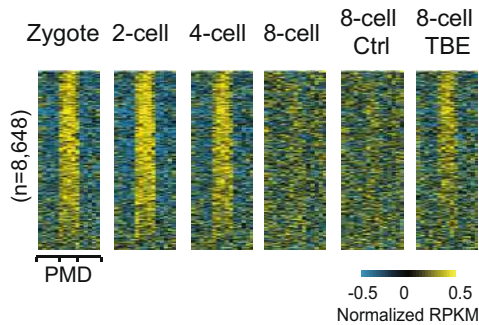
c Distal ATAC-seq peaks



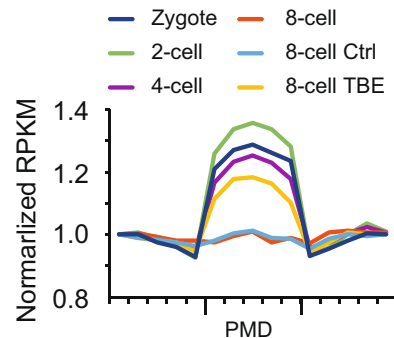
d ATAC-seq peaks in oocyte PMD



e ATAC-seq enrichment in oocyte PMD

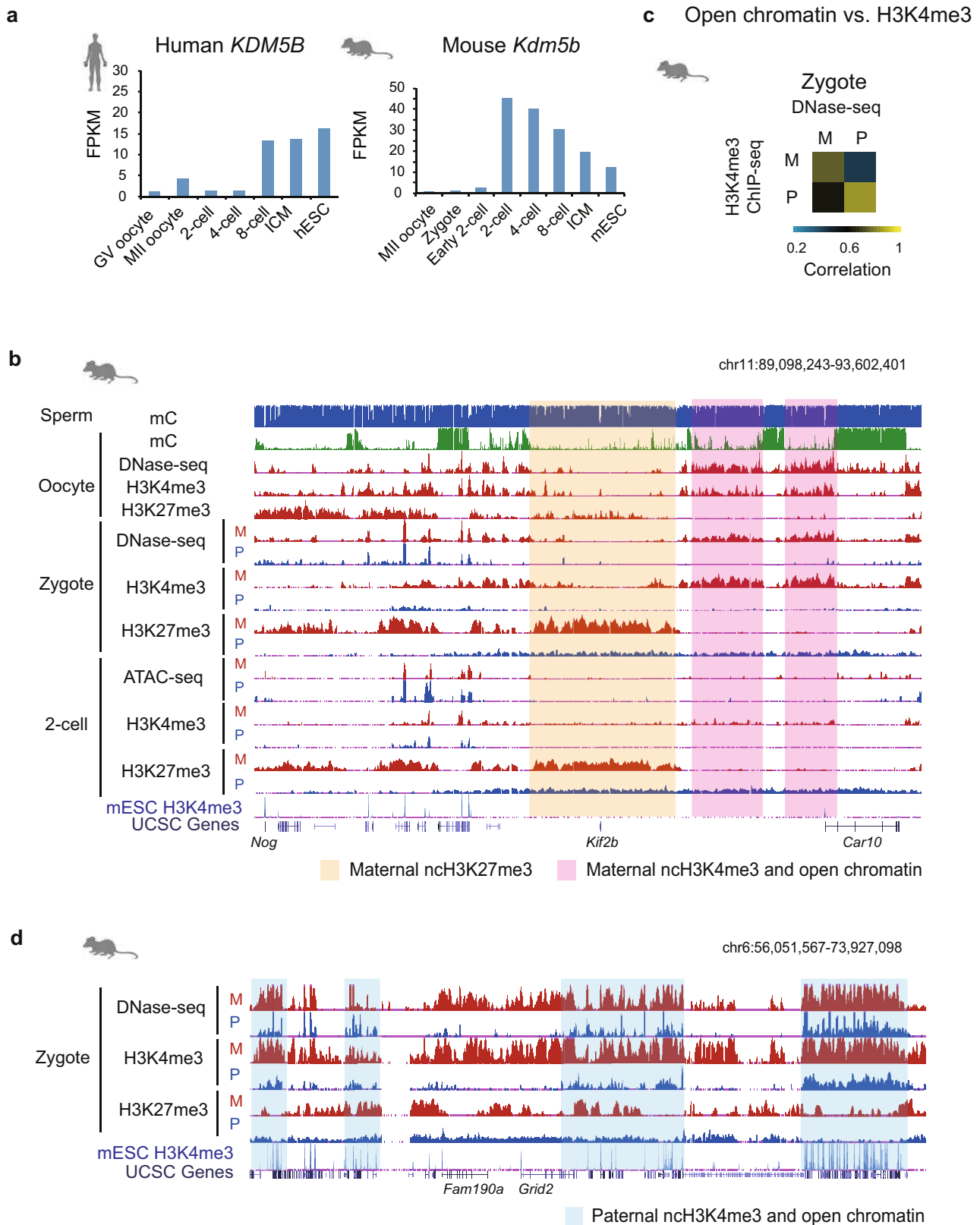


f Average ATAC-seq enrichment



Extended Data Fig. 8 | Characterization of two-cell distal open chromatin in human embryos. **a**, Heat maps showing the ATAC-seq signals and CpG density for promoter and distal ATAC-seq peaks at the two-cell stage. **b**, Top, Venn diagram showing the overlap between two- and eight-cell distal ATAC-seq peaks. Bottom, motifs identified in two- and eight-cell embryo shared distal peaks as well as peaks specific for each stage are also shown. **c**, The enrichment of human ES cell H3K4me1 and H3K27ac marks⁴⁷ around two-cell-specific distal peaks or two-to-eight-cell shared distal peaks is shown. A random set of peaks that match the lengths of individual two-cell-specific ATAC-seq peaks was used as a

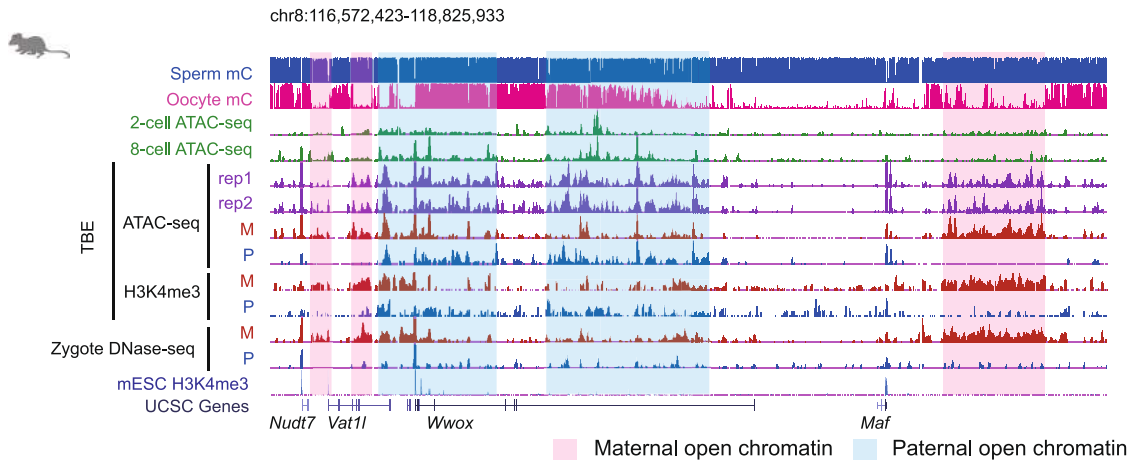
control. The upstream and downstream regions are $2 \times$ peak lengths away from peak boundary. **d**, Bar chart showing the percentages of ATAC-seq peaks that overlap with oocyte PMDs for total peaks, peaks shared by various stages, or peaks specific for each stage. **e**, Heat maps showing the enrichment of ATAC-seq around the human oocyte PMDs in human early embryos and TBEs. The upstream and downstream regions are $1 \times$ PMD length away from the PMD boundary. **f**, The average enrichment of ATAC-seq around the oocyte PMDs in human early embryos and TBEs. The upstream and downstream regions are $1 \times$ PMD length away from the PMD boundary.



Extended Data Fig. 9 | Accessible chromatin state in mouse oocytes and pre-ZGA embryos. **a**, The expression of human *KDM5B* and mouse *Kdm5b* in oocytes, early embryos and ES cells. **b**, UCSC browser view showing DNA methylation levels (mC) in mouse sperm and oocyte, as well as open chromatin (DNase-seq or ATAC-seq) and H3K4me3 and H3K27me3 enrichment in mouse oocytes, zygotes and two-cell embryos. Mouse ES cell H3K4me3 signals are also shown to mark the promoter

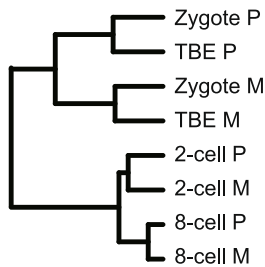
regions. **c**, Heat map showing the Spearman correlation between allelic DNase-seq and H3K4me3 signals in zygotes. M, maternal; P, paternal. **d**, UCSC browser view showing allelic DNase-seq, H3K4me3 and H3K27me3 enrichment in the mouse zygote. The mouse ES cell H3K4me3 signal is also shown. Regions showing paternal open chromatin and H3K4me3 in zygotes are shaded.

a



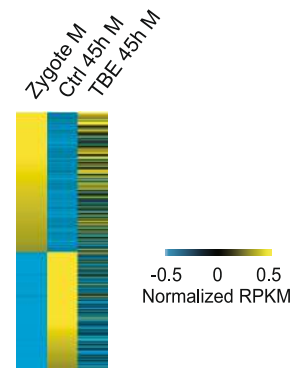
b

Hierarchical clustering of open chromatin



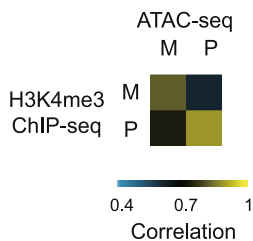
c

Differential peaks (pre- vs. post ZGA)

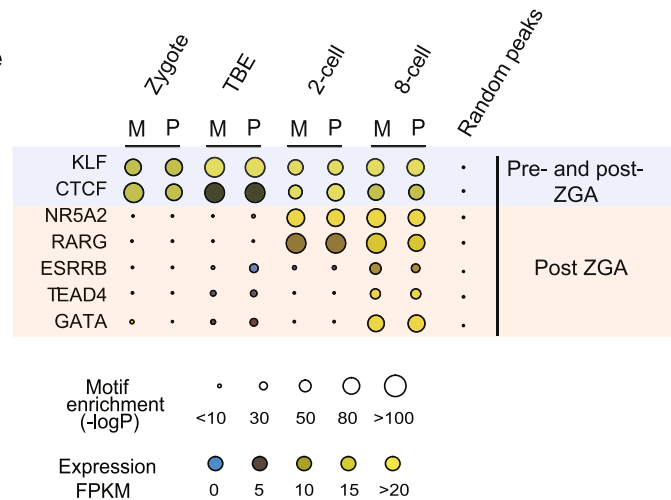


d

Open chromatin vs. H3K4me3 (TBE)



e



Extended Data Fig. 10 | Accessible chromatin state in mouse TBES. a, UCSC browser view showing DNA methylation levels in mouse sperm and oocyte, and ATAC-seq signals in normal mouse embryos as well as allelic ATAC-seq and H3K4me3 enrichment in TBE samples. Mouse ES cell H3K4me3 signals are also shown to mark the promoter regions. b, Hierarchical clustering results showing the relationships of allelic accessible chromatin states between zygotes, TBES, and two- and eight-cell embryos in mouse. c, Heat map showing open chromatin regions that are unique to zygotes (DNase-seq) or 45h control embryos (ATAC-seq). The ATAC-seq enrichment in TBE samples is then matched and shown. d, Heat map showing the Spearman correlation

between allelic ATAC-seq and H3K4me3 signals in TBES. e, Transcription factor motifs identified from distal DNase-seq and allelic distal ATAC-seq peaks are shown. Motifs shared by pre-ZGA and post-ZGA stages or are specific for post-ZGA stages are noted. For transcription factors that have multiple family members with similar motifs (KLF and GATA), the highest expression and motif enrichment among all family members at each stage are shown. A random set of peaks that match the lengths and number of zygote maternal peaks was used as a control. It is worth noting that the RNA levels of CTCF appear to decline in TBES, presumably owing to RNA degradation during extended transcription inhibition.

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

At least 30 cells are used for each sample. Our analysis (Extended Data Figure 1) show that miniATAC-seq using 30 cells can provide robust results.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

Our findings are well reproduced in replicates

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

All the gametes and embryos were collected from the volunteers who were between 25 and 30 years old, with chromosome karyotype normal and no family heredity case history, contagion case and smoke history.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding is not relevant to this study as most of our data do not involve control vs. treatment. For other experiments, variations were controlled through replicates.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Bowtie (version 2.2.2);Tophat (version 2.4.0) ; Cufflinks (version 2.2.1) ;BSMAP(v2.74);HOMER; MACS14;Bedtools; Samtools;DAVID;GREAT;picard-tools-1.84;R

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

H1 hESC; H7 hESC

b. Describe the method of cell line authentication used.

H1 hESCs were obtained from WiCell Institute and authenticated by cytogenetic test and genomic DNA sequencing; H7 hESCs were obtained from WiCell Institute and authenticated by genomic DNA sequencing.

c. Report whether the cell lines were tested for mycoplasma contamination.

PCR test were performed to confirm cell are negative for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No misidentified cell lines maintained were used

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

5 to 6 week-old C57BL/6N females (Vital River) and PWK/PhJ males (Jackson Laboratory) mice were used

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All the gametes and embryos were collected from the volunteers who were between 25 and 30 years old, with chromosome karyotype normal and no family heredity case history, contagion case and smoke history